

Exhibit A
Structural Domain Analysis of SEQ ID NO: 2

INTERPRO

InterPro is a database of protein families, domains and functional sites in which identifiable features found in known proteins can be applied to unknown protein sequences. <http://www.ebi.ac.uk/interpro/>

noIPR	<u>PD472437</u>	sp_Q9VFP6_DROME_Q9VFP6	unintegrated
unintegrated	<u>PD004810</u>	sp_Q8VCL5_MOUSE_Q8VCL5	
	<u>PD513011</u>	sp_Q9BYT1_HUMAN_Q9BYT1	
	<u>PD000916</u>	sp_Q9BYT1_HUMAN_Q9BYT1	
	<u>PD041621</u>	sp_Q95X51_CAEEL_Q95X51	
	<u>PD434467</u>	sp_Q9BYT1_HUMAN_Q9BYT1	
	<u>PD001152</u>	sp_Q9BYT1_HUMAN_Q9BYT1	
	<u>PD394380</u>	sp_Q9DA66_MOUSE_Q9DA66	
	<u>PD508204</u>	sp_Q9BYT1_HUMAN_Q9BYT1	
<u>IPR005829</u>			
Family	<u>PS00217</u>	SUGAR_TRANSPORT_2	Sugar transporter superfamily
<u>IPR005828</u>			
Family	<u>PF00083</u>	PF00083	General substrate transporter

Pfam

Pfam is a large collection of multiple sequence alignments and hidden Markov models covering many common protein domains and families. For each family in Pfam you can: Look at multiple alignments; View protein domain architectures; Examine species distribution; Follow links to other databases; View known protein structures.
<http://www.sanger.ac.uk/Software/Pfam/index.shtml>

Model	Seq-from	HMM-from	HMM-to	Score	E-value	Alignment	Description
<u>!! sugar_tr</u>	23	429	1	487	-119.5	0.0007	global Sugar (and other) transporter
sugar_tr:	domain 1 of 1, from 23 to 429: score -119.5, E = 0.0007						
	*->valvaalgggf1fGyDtgviggflalidflfrfglltssgalas1vg						
	+ a++G++l G + +++ +++i + +++ a						
sequence	23	--CQAWTGTLLLGTCLLYCARSSMPICTVSMSQDFGWNKKEA----	62				
sequence	63	ystvltglvvsiifflGrliGslfaGklgdrfGRkksllial...vlfv...g+v s+ff G + ++++G+lgdr+G k +l+++ + + ++	106				
sequence	107	GallsgaapgytTiGlwafyllivGrvlvGlgvGgasvlvPmYisEiAPk	149				
		11 +++ + + + R+l+G1 G+ + + ++S+ +					
		+ R++ s+ + ++G l++ +g l + ++ W + +++					
		a1RGalgslyqlaitiGilvAaiiglglnktnndsalnswgWRiplglql					

sequence 150 SERAFTYSIVGAGSQFGTLLTGAVGSLLEWYG-----WQSIFYFSG 191
 vpalllligllf1PESPRwLvekgkleeArevLaklrgvedvdqeiqueek
 ++ 1++++ + R+L++ ++1 A vLa+r +
 sequence 192 GLT---LLWVWYVY---RYLLSEKDLILALGVLAQSR-----P 223
 sequence 224 VSRHRSRVP-----WRRLFRKPA-----VWA AVVS QLSA-ACSFFIL 258
 aeleatvseekagkaswgelfrgrtpkyrqrl1mgvmlqafqQltGiNai
 ++ + v+ w+ lfr + ++ + v+ q+ + + +
 sequence 259 LSWLPTFFEETFPDAK--GWIFNVVPWLVAIPASLFS-GFLSDHlinqgy 305
 ..fGRRpl11.1Gaagmaicflilgasvival11lnkpkdpsksaagiva
 + ++ R+l+ .+G+ . ++++1 lg +++++ a
 sequence 306 raITVRKLMQgMGLGLSSVFALCLG--HTSSFCESV-----VFA 342
 ivfillfiaifFalgwGpipwvilsElFPtkvRskalalataanwlanfii
 + i l + ++ g+ v ++++1 P ++ + +++a a+ la+++
 sequence 343 SASIGLQTFNHS-GIS---VNIQDLAP-SCAGFLFGVANTAGALAGVVG 386
 gflfpwyitgaiglalggvvflvfagllvlfilfvffffvPETkGrtLEeie
 1 y+++ g + f++ a++ l+ + f+v G ++
 sequence 387 VCLGGYLMETTG--SWTCLFNLVAIISNLGL--CTFLVF---G---QAQR 426
 elf<-*
 +++
 sequence 427 VDL 429

ProtComp

<http://www.hgmp.mrc.ac.uk/GenomeWeb/prot-anal.html>
 ProtComp Version 5. Identifying sub-cellular location (Animals&Fungi)

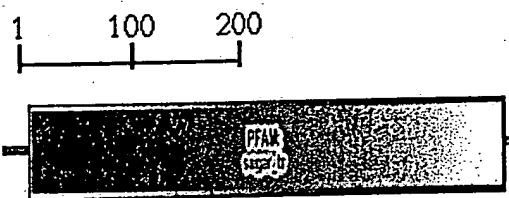
Seq name: sequence 436
 Significant similarity in Location DB - Location: Plasma membrane
 Database sequence: AC=Q9BYT1 Location: Plasma membrane DE BA305P22.2.1 (Novel protein, isoform 1).
 Score=21855, Sequence length=430, Alignment length=422
 Predicted by Neural Nets - Plasma membrane with score 2.9
 ***** Transmembrane segments are found: .+166:179+.+275:295-.+399:412-.
 ***** Potential GPI-anchor in position 414 is found
 Integral Prediction of protein location: Plasma membrane with score 7.8
 Location weights: LocDB / PotLocDB / Neural Nets / Integral

Nuclear	0.0 /	0.0 /	0.73 /	0.73
Plasma membrane	21855.0 /	0.0 /	2.91 /	7.77
Extracellular	0.0 /	0.0 /	0.72 /	0.72
Cytoplasmic	0.0 /	0.0 /	0.66 /	0.66
Mitochondrial	0.0 /	0.0 /	0.69 /	0.69
Endoplasm. retic.	0.0 /	0.0 /	0.71 /	0.71
Peroxisomal	0.0 /	0.0 /	0.38 /	0.38
Lysosomal	0.0 /	0.0 /	0.26 /	0.26
Golgi	0.0 /	0.0 /	0.24 /	0.24

SMART ANALYSIS

SMART (a Simple Modular Architecture Research Tool) allows the identification and annotation of genetically mobile domains and the analysis of domain architectures. More than 500 domain families found in signalling, extracellular and chromatin-associated proteins are detectable. These domains are extensively annotated with respect to phyletic distributions, functional class, tertiary structures and functionally important residues. Each domain found in a non-redundant protein database as well as search parameters and taxonomic information are stored in a relational database system. User interfaces to this database allow searches for proteins containing specific combinations of domains in defined taxa. **For all the details, please refer to the publications on SMART:**
<http://smart.embl-heidelberg.de/>

Domains within the query sequence sequence of 436 residues



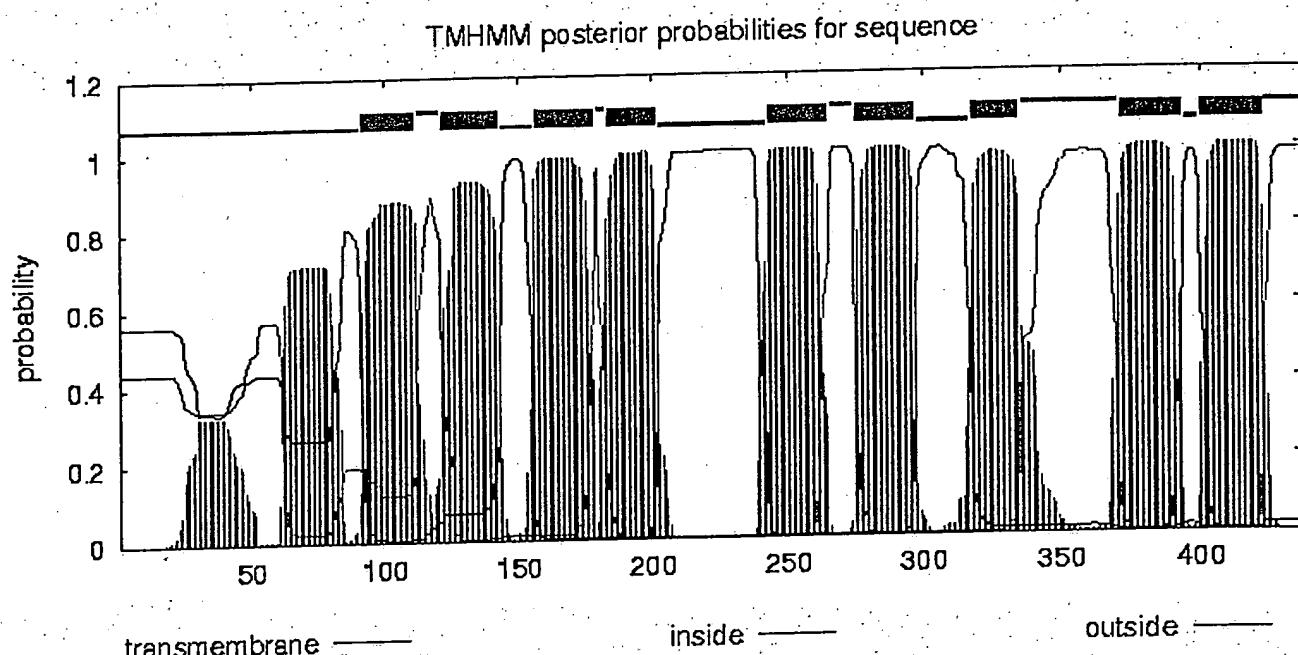
Smart Analysis tells of a sugar_tr domain
Accession number: PF00083
Sugar (and other) transporter

TMHMM

Prediction of transmembrane helices in proteins
<http://www.cbs.dtu.dk/services/TMHMM/>

```
# sequence Length: 436
# sequence Number of predicted TMHs: 9
# sequence Exp number of AAs in TMHs: 215.57889
# sequence Exp number, first 60 AAs: 6.90461
# sequence Total prob of N-in: 0.55992
sequence TMHMM2.0 inside 1 91
sequence TMHMM2.0 TMhelix 92 111
sequence TMHMM2.0 outside 112 120
sequence TMHMM2.0 TMhelix 121 143
sequence TMHMM2.0 inside 144 155
sequence TMHMM2.0 TMhelix 156 178
sequence TMHMM2.0 outside 179 182
sequence TMHMM2.0 TMhelix 183 202
sequence TMHMM2.0 inside 203 242
sequence TMHMM2.0 TMhelix 243 265
sequence TMHMM2.0 outside 266 274
sequence TMHMM2.0 TMhelix 275 297
sequence TMHMM2.0 inside 298 317
sequence TMHMM2.0 TMhelix 318 335
```

sequence	TMHMM2.0	outside	336	370
sequence	TMHMM2.0	TMhelix	371	393
sequence	TMHMM2.0	inside	394	399
sequence	TMHMM2.0	TMhelix	400	422
sequence	TMHMM2.0	outside	423	436



ProDom

ProDom is a comprehensive set of protein domain families automatically generated from the SWISS-PROT and TrEMBL sequence databases. **Nucl. Acids. Res.** Corpet et al. 28 (1): 267.
<http://prodom.toulouse.inra.fr/prodom/2002.1/html/home.php>

HSP Results

Warning: Original output has been filtered to yield non-redundant similarities
 BLASTP 2.2.1 [Apr-13-2001]

Reference: Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schaffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", *Nucleic Acids Res.* 25:3389-3402.

Query=

(436 letters)

Database: ProDom 2002.1 Jan2003 multiple alignments
 1,619,602 sequences; 167,025,341 total letters

Searching.....done

ProDom domains producing High-scoring Segment Pairs:

Position	ProDom domain	Score	E value
13-86	#PD004810	325	4e-31
38-104	#PD003131	82	0.007
38-104	#PD523332	86	0.002
38-107	#PD535883	107	8e-06
45-95	#PD413016	89	0.001
48-108	#PD543895	87	0.002
55-92	#PD063885	89	0.001
55-108	#PD000036	95	2e-04
67-108	#PD000082	109	5e-06
87-132	#PD513011	165	2e-12
125-191	#PD078712	87	0.002
133-192	#PD000916	310	2e-29
207-255	#PD434467	244	1e-21
255-318	#PD413016	113	2e-06
256-325	#PD001152	369	4e-36
279-346	#PD394380	183	1e-14
327-371	#PD286146	91	6e-04
347-427	#PD508204	311	2e-29

>PD001152 (Closest domain: Q9BYT1_HUMAN 250-319)

Number of domains in family:

Commentary (automatic):

TRANSPORTER INORGANIC RENAL SODIUM NA-DEPENDENT

Length = 70

Score = 369. (146 bits), Expect = 4e-36

Identities = 70/70 (100%), Positives = 70/70 (100%)

Query: 256 FILLSWLPTFFFEETFPDAKGWIFNVVPWLVAIPASLFGFLSDHLINQGYRAITVRKLMQ 315
FILLSWLPTFFFEETFPDAKGWIFNVVPWLVAIPASLFGFLSDHLINQGYRAITVRKLMQ

Sbjct: 250 FILLSWLPTFFFEETFPDAKGWIFNVVPWLVAIPASLFGFLSDHLINQGYRAITVRKLMQ 309

Query: 316 GMGLGLSSVF 325

GMGLGLSSVF

Sbjct: 310 GMGLGLSSVF 319

>PD004810 (Closest domain: Q8VCL5_MOUSE 11-89)

Number of domains in family:

Commentary (automatic):

GLYCOPROTEIN CHROMOSOME NA-DEPENDENT III SYMPORT SODIUM

Length = 79

Score = 325 (129 bits), Expect = 4e-31

Identities = 57/74 (77%), Positives = 61/74 (82%)

Query: 13 AGDTQWSRPECQAWXXXXXXXXXXXXCARSSMPICTVMSQDFGWNKKEAGIVLSSFFWG 72
A DT+WSRPECQAW YCAR +MP+CTV+MSQDFGWNKKEAGIVLSSFFWG

Sbjct: 16 AEDTRWSRPECQAWTGILLGTCLLYCARVTMPVCTVAMSQDFGWNKKEAGIVLSSFFWG 75

Query: 73 YCLTQVVGGHLGDR 86

YCLTQVVGGHLGDR

Sbjct: 76 YCLTQVVGGHLGDR 89

>PD508204 (Closest domain: Q9BYT1_HUMAN 341-421)

Number of domains in family:

Commentary (automatic):

NA-DEPENDENT BAC NOVEL SIMILAR THALIANA ARABIDOPSIS

Length = 81

Score = 311 (124 bits), Expect = 2e-29

Identities = 62/81 (76%), Positives = 62/81 (76%)

Query: 347 GLQTFNHSGISVNIQDLAPSCAGFLXXXXXXXXXXXXXXYLMETTGSWTCLFNL 406
GLQTFNHSGISVNIQDLAPSCAGFLF YLMETTGSWTCLFNL

Sbjct: 341 GLQTFNHSGISVNIQDLAPSCAGFLFGVANTAGALAGVVGVCLGGYLMETTGSWTCLFNL 400

Query: 407 VAIISNLGLCTFLVFGQAQRV 427

VAIISNLGLCTFLVFGQAQRV

Sbjct: 401 VAIISNLGLCTFLVFGQAQRV 421

>PD000916 (Closest domain: Q9BYT1_HUMAN 127-199)

Number of domains in family:

Commentary (automatic):

RESISTANCE MEMBRANE PROBABLE MULTIDRUG FAMILY

Length = 73

Score = 310 (124 bits), Expect = 2e-29

Identities = 60/60 (100%), Positives = 60/60 (100%)

Query: 133 GVYFPALTSSLSQLVRESERAFTYSIVGAGSQFGTLTGAVGSLLEWYGWQSIFYFSGG 192
GVYFPALTSSLSQLVRESERAFTYSIVGAGSQFGTLTGAVGSLLEWYGWQSIFYFSGG

Sbjct: 127 GVYFPALTSSLSQLVRESERAFTYSIVGAGSQFGTLTGAVGSLLEWYGWQSIFYFSGG 186

>PD434467 (Closest domain: Q9BYT1_HUMAN 200-249)

Number of domains in family: 1

Commentary (automatic):

Length = 50

Score = 244 (98.6 bits), Expect = 1e-21

Identities = 48/49 (97%), Positives = 49/49 (99%)

Query: 207 SEKDLILALGVLAQSRPVSRRHVPWRRLFRKPAVWAUVSQLSAACSF 255
SEKDLILALGVLAQSRPVSRRH+RVPWRRLFRKPAVWAUVSQLSAACSF

Sbjct: 201 SEKDLILALGVLAQSRPVSRRHVPWRRLFRKPAVWAUVSQLSAACSF 249

>PD394380 (Closest domain: Q9DA66_MOUSE 1-99)

Number of domains in family: 1

Commentary (automatic):

Length = 99

Score = 183 (75.1 bits), Expect = 1e-14

Identities = 41/74 (55%), Positives = 54/74 (72%), Gaps = 7/74 (9%)

Query: 279 NVV PWLVAIPASLFGFLSDHLIN-----QGYRAITVRKLMQGMGLGLSSVFALCLGHT 332
N++P ++ + L S L+ HL+ QGYR ITVRK MQ MGLGLSS+FALCLGHT

Sbjct: 27 NLLPVVLCL-LLLHSTLLAAHLLQGDLPQLQGYRVITVRKFMQVMGLGLSSIFALCLGHT 85

Query: 333 SSFCESVVFASASI 346

+SF +++++FASASI

Sbjct: 86 TSFLKAMIFASASI 99

>PD513011 (Closest domain: Q9BYT1_HUMAN 81-126)

Number of domains in family: 1

Commentary (automatic):

Length = 46

Score = 165 (68.2 bits), Expect = 2e-12

Identities = 35/46 (76%), Positives = 35/46 (76%)

Query: 87 IGGEKVILLSASAWSITAVTPXXXXXXXXXXFMTFSRILMGLLQ 132

IGGEKVILLSASAWSITAVTP FMTFSRILMGLLQ

Sbjct: 81 IGGEKVILLSASAWSITAVTPLAHLSAHLAFMTFSRILMGLLQ 126

>PD413016 (Closest domain: Q8W4P5_ARATH 352-432)

Number of domains in family: 895

Commentary (automatic):

MULTIDRUG PROBABLE EFFLUX PERMEASE

Length = 81

Score = 113 (48.1 bits), Expect = 2e-06

Identities = 24/67 (35%), Positives = 35/67 (51%), Gaps = 4/67 (5%)

Query: 255 FFILLSWLPTFFETFP---DAKGWIFNVVPWLVAIPASLFSGFLSDHLINQGYRAITVR 311
FF++LSW+P +F + W F+ VPW + +GF SD LI +G R

Sbjct: 353 FFVILSWMPIYFNSVYHVNLKQAAW-FSAVPWSMMAFTGYIAGFWSDLIRRGTSITLTR 411

Query: 312 KLMQGMG 318

K+MQ +G

Sbjct: 412 KIMQSIG 418

>PD000082 (Closest domain: Q9SH82_ARATH 142-197)

Number of domains in family:

Commentary (automatic):

RESISTANCE MEMBRANE PROBABLE FAMILY MULTIDRUG

Length = 56

Score = 109 (46.6 bits), Expect = 5e-06

Identities = 19/42 (45%), Positives = 26/42 (61%)

Query: 67 SSFWGYCLTVVGHHLGDRIGGEKVILLSASAWSITAVTP 108

SSF WGY + V+GG L DR GG++V+ + W T +TP

Sbjct: 142 SSFLWGYIFSSVIGGALVDRYGGKRVLAWSLATLLTP 183

>PD535883 (Closest domain: Q8YJH9_BRUME 1-144)

Number of domains in family: 1

Commentary (automatic):

Length = 144

Score = 107 (45.8 bits), Expect = 8e-06

Identities = 21/70 (30%), Positives = 42/70 (60%), Gaps = 1/70 (1%)

Query: 38 YCARSSMPCTVSMSQDFGWNKKEAGIVLSSFFWGYCLTVVGHHLGDRIGGEKVILLSA 97
Y R ++ + + ++G+N+ + G +L F +GY ++GG L D++G K+ +++

Sbjct: 49 YIDRGAISYASEQIIIGEYGFNRADWGSMLGYFGYGYMGAILGGTLSDKLGARKLWIAG 108

Query: 98 SAWGSITAVT 107

+AW SI AV+

Sbjct: 109 TAW-SIVAVS 117

>PD000036 (Closest domain: Q9V905_DROME 63-130)

Number of domains in family:

Commentary (automatic):

SODIUM-DEPENDENT CARRIER SODIUM-PHOSPHATE SODIUM FAMILY

Length = 68

Score = 95 (41.2 bits), Expect = 2e-04

Identities = 16/54 (29%), Positives = 30/54 (54%)

Query: 55 FGWNKKEAGIVLSSFWGYCLTQVVGHLGDRIGGEKVILLSASAWGSITAVTP 108

+ W + + ++L ++F+GY +T + G L + +G V S G +TA+TP

Sbjct: 63 YNWTQSDQALLLGAYFYGYMITSLPAGTLAEMLGARNVAGYSCLVAGILTALTP 116

>PD286146 (Closest domain: Q9SH82_ARATH 407-561)

Number of domains in family: 1

Commentary (automatic):

Length = 155

Score = 91 (39.7 bits), Expect = 6e-04

Identities = 18/45 (40%), Positives = 28/45 (62%)

Query: 327 LCLGHTSSFCESVVFASASIGLQTFNHSGISVNIDLAPSCAGFL 371

LCL S + VF + ++ L +F+ +G +N+QD+AP AGFL

Sbjct: 411 LCLNFAKSPSCAAVFMTIALSLSFSQAGFLNMQDIAPQYAGFL 455

>PD413016 (Closest domain: Q99TA7_STAAM 17-104)

Number of domains in family: 895

Commentary (automatic):

MULTIDRUG PROBABLE EFFLUX PERMEASE

Length = 88

Score = 89 (38.9 bits), Expect = 0.001

Identities = 19/51 (37%), Positives = 31/51 (60%)

Query: 45 PICTVSMSQDFGWNKKEAGIVLSSFWGYCLTQVVGHLGDRIGGEKVILL 95

P+ T+ M Q+ G+ AG+VL +G + ++GG L D++GG K IL+

Sbjct: 26 PLNTIYMKQELGKSLTVAGLVLMINSFGMVIGNLLGGSLFDLGGYKTIIL 76

>PD063885 (Closest domain: Q9V763_DROME 1-161)

Number of domains in family: 2

Commentary (automatic):

COTRANSPORTER

Length = 161

Score = 89 (38.9 bits), Expect = 0.001

Identities = 16/38 (42%), Positives = 23/38 (60%)

Query: 55 FGWNKKEAGIVLSSFWGYCLTQVVGHLGDRIGGEKV 92

F WN+K+ G +L SFFW + Q+ GG L + G + V

Sbjct: 83 FHWNEKQQGALLGSFFWAHWTLQIPGGILATKYGTKLV 120

>PD078712 (Closest domain: Q23063_CAEEL 5-202)

Number of domains in family: 3

Commentary (automatic):

Length = 198
Score = 87 (38.1 bits), Expect = 0.002
Identities = 22/69 (31%), Positives = 34/69 (48%), Gaps = 2/69 (2%)
Query: 125 RILMGLLQGVYFPALTSLLSQKVRESERAFTYSIVGAGSQFGTLLTGAVSLLL--EWYG 182
R G Q L+ + ESE +F +SI+ A SQFG L T +G + ++G
Sbjct: 118 RFFAGFAQASQLHFTNDLVLWRTPESAEASFFFSSIMLATSQFGPLFTMILGGEMCSSSFFG 177
Query: 183 WQSIFYFSG 191
W++ +Y G
Sbjct: 178 WEATYYILG 186

>PD543895 (Closest domain: Q8ZR98_SALTY 217-325)

Number of domains in family: 8

Commentary (automatic):

TRANSMEMBRANE MEMBRANE ANTIBIOTIC

Length = 109
Score = 87 (38.1 bits), Expect = 0.002
Identities = 22/61 (36%), Positives = 32/61 (52%), Gaps = 6/61 (9%)
Query: 48 TVSMSQDFGWNKKEAGIVLSSFFWGYCLTVVGGHLDRIGGEKVILLSASAWGSITAVT 107
T + Q FG + + A + L +F + V+GG +GD+IG + VI WGSI V
Sbjct: 251 TFYLMQKFGLSIQNAQLHLFAFLFAVAAGTVIGGPVGDKIGRKYVI-----WGSILGVA 304
Query: 108 P 108
P
Sbjct: 305 P 305

>PD523332 (Closest domain: Q8ZK69_SALTY 1-107)

Number of domains in family: 10

Commentary (automatic):

PERMEASE PROBABLE 2-KETOGLUCONATE INTEGRAL

Length = 107
Score = 86 (37.7 bits), Expect = 0.002
Identities = 18/67 (26%), Positives = 32/67 (46%)
Query: 38 YCARSSMPICTVSMSQDFGWNKKEAGIVLSSFFWGYCLTVVGGHLDRIGGEKVILLSA 97
Y RS++ + +++ D + A IVL+ F GY + + GG R +K+++L
Sbjct: 22 YLDRSNLSVTLPITHDLNIDGATASIVLTIFLYAFSNIFGGVFTQRYPKIVILMV 81
Query: 98 SAWGSIT 104
W T
Sbjct: 82 LIWSIAT 88

>PD003131 (Closest domain: Q9RPP3_BURCE 19-130)

Number of domains in family:

Commentary (automatic):

PLASMID PROBABLE 4-HYDROXYPHENYLACETATE MFS PHTHALATE

Length = 112
Score = 82 (36.2 bits), Expect = 0.007
Identities = 20/67 (29%), Positives = 31/67 (45%)
Query: 38 YCARSSMPICTVSMSQDFGWNKKEAGIVLSSFFWGYCLTVVGGHLDRIGGEKVILLSA 97
Y R ++ + + D G + G+ +S FF GY L +V L RIG K +
Sbjct: 45 YLDRVNVSAQLQLKHDLGLSDAAAYGLGVSLFFIGYILLEVPSTLLRRIGARKTVTRIM 104

Query: 98 SAWGSIT 104

WG+I+

Sbjct: 105 LLWGAIS 111

Parameters:

Database: ProDom 2002.1 Jan2003 multiple alignments

Number of letters in database: 167,025,341

Number of sequences in database: 1,619,602

Lambda K H
0.325 0.138 0.441

Gapped

Lambda K H
0.267 0.0410 0.140